AD-A045 410 CALIFORNIA UNIV. DAVIS DEPT OF MATHEMATICS F/G 12/1

MAXIMUM LIKELIHOOD ESTIMATION FOR BINOMIALLY DISTRIBUTED SIGNAL--ETC(U)

JUL 77 F J SAMANIEGO AFOSR-77-3180

UNCLASSIFIED TR-3 AFOSR-TR-77-0947 NL

1 OF 1

ADA045410

END

DATE FILMED

11-77

DDC

AFOSR-TR-77-0947

TECHNICAL REPORT SERIES ON SIGNAL PLUS NOISE MODELS

Department of Mathematics
University of California
Davis, California

ROUTING AND TRANSMITTAL SLIP

TO (Name, office symbol or location)		ACTION
1 DDC/TCA/Mrs Crumbacker	INITIALS	CIRCULATE
	DATE	COORDINATION
2	INITIALS	FILE
	DATE	INFORMATION
3	INITIALS	NOTE AND RETURN
	DATE	PER CON-VERSATION
4	INITIALS	SEE ME
	DATE	SIGNATURE

REMARKS

Attached is a revised version of AFOSR-TR-77-0947 which had ADA043227. Supply

A045410

Do NOT use this form as a RECORD of approvals, concurrences, disapprovals, clearances, and similar actions

FROM (Name, office symbol or location)	DATE
AFOSR/XOPD/[illegible]	13 Oct 77
	PHONE [illegible]-4912

OPTIONAL FORM 41
AUGUST 1967
GSA FPMR (41 CFR) 100-11.206

GPO 043-16-[illegible] 410-013 5041-101

ACCESSION for
NTIS White Section
DDC Buff Section
UNANNOUNCED
JUSTIFICATION
BY
DISTRIBUTION/AVAILABILITY CODES
Dist AVAIL. and/or SPECIAL
A

(4)

Maximum Likelihood Estimation for Binomially Distributed Signals in Discrete Noise

Francisco J. Samaniego

Technical Report No. 3

July, 1977

Department of Mathematics
University of California
Davis, California

This research was supported in part by the Air Force Office of Scientific Research under contract AFOSR-77-3180.

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER AFOSR-TR-77-0947	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) MAXIMUM LIKELIHOOD ESTIMATION FOR BINOMIALLY DISTRIBUTED SIGNALS IN DISCRETE NOISE.		5. TYPE OF REPORT & PERIOD COVERED Interim technical rept.,
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Francisco J. Samaniego		8. CONTRACT OR GRANT NUMBER(s) AFOSR-77-3180
9. PERFORMING ORGANIZATION NAME AND ADDRESS University of California Department of Mathematics Davis, CA 95616		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 61102F 2304/A5
11. CONTROLLING OFFICE NAME AND ADDRESS Air Force Office of Scientific Research/NM Bolling AFB, Washington, DC 20332		12. REPORT DATE July 1977
		13. NUMBER OF PAGES 23
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office) 25p. TR-3		15. SECURITY CLASS. (of this report) UNCLASSIFIED
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

Supersedes AO43227

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

maximum likelihood estimation, convolution, parametric monotone ratio

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Let X be a discrete variable distributed like the sum of independent variables Y and Z, where the signal Y is binomially distributed and the noise variable Z is a nonnegative integer valued variable whose distribution does not depend on the binomial parameter p. The family of convoluted binomial distributions, that is, distributions of variables such as X, are characterized in terms of a system of differential equations satisfied by their

next page

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

401 891

20. Abstract

probability mass functions. A monotonicity property for probability ratios of certain convoluted binomial distributions is noted, and the maximum likelihood estimate of the binomial parameter is shown to be easily obtained for these distributions using the characterizing property established for their mass functions. Results are applicable to models for binomial signals in noise in which the noise distribution is known or can be estimated from an auxiliary experiment.

SUMMARY

Let X be a discrete variable distributed like the sum of independent variables Y and Z, where the signal Y is binomially distributed and the noise variable Z is a nonnegative integer valued variable whose distribution does not depend on the binomial parameter p. The family of convoluted binomial distributions, that is, distributions of variables such as X, are characterized in terms of a system of differential equations satisfied by their probability mass functions. A monotonicity property for probability ratios of certain convoluted binomial distributions is noted, and the maximum likelihood estimate of the binomial parameter is shown to be easily obtained for these distributions using the characterizing property established for their mass functions. Results are applicable to models for binomial signals in noise in which the noise distribution is known or can be estimated from an auxiliary experiment.

I. INTRODUCTION

Consider a random variable X distributed as the sum of independent components Y and Z. Many situations in which a variable such as X arises have the characteristic that one of the components of X, say Y, is of particular importance to the investigator, while the other component Z is of lesser or no interest. Typical of such scenarios is the process of making observations with a Gieger counter, where the observed count is the sum of counts due to the presence of radioactive material and counts due to noise or static. We refer to X as a signal plus noise variable, and identify the component of X of interest in a particular study as the signal variable.

Estimation for signal plus noise models has been pursued by a number of authors. Research on such models may be classified into two categories -- those which postulate that the noise distribution is known and those that do not. Among investigations of the first type is the study by Gaffey (1959), in which a consistent estimator of the signal distribution was obtained with the noise variable modeled as a normal variable with both parameters known. Samaniego (1976) studied maximum likelihood estimation for Poisson signals in discrete noise with known distribution. He showed that the maximum likelihood estimate of the Poisson parameter could be obtained for a broad class of models via a characterization result established for convoluted Poisson distributions. The assumption that the noise distribution is known appears quite restrictive, but is realizable in experimental situations in which the noise variable may be observed alone. In such situations, the empirical distribution of the noise variable may serve to approximate the true noise distribution to any desired degree of accuracy. Estimation for models in which only the functional form of the noise distribution is known has been studied by Sclove and Van Ryzin (1969), who obtained method of moments estimators and their asymptotic variances for a variety of models.

We treat in this paper maximum likelihood estimation for binomial signals in discrete noise with known distribution. For the special case of the Binomial-Poisson model, moment estimators for the parameters appear in Sclove and Van Ryzin (1969), while the maximum likelihood estimate of the Poisson parameter when the binomial parameter is known may be obtained from work in Samaniego (1976). The complementary problem of estimating the binomial parameter by maximum likelihood methods when the Poisson parameter is known is treated as an example in Section III of this paper.

Binomial signals in noise arise in a variety of contexts. Examples abound in the literature on statistical communication theory. Woodward (1953) discusses a communication system in which the received signal is the sum of independent Bernoulli variables. Examples of recent work on discrete signal detection include that of Greenstein (1974) on block coding of binary signals and work by Crochiere et al. (1976) on digital coding of speech. Another source of data well modelled as binomial signals in discrete noise is the area of acceptance sampling. The number of defectives found in a sample from a large population is usually modelled as a binomial random variable. Inspection policies which result in systematically overcounting or undercounting the number of defectives in the sample give rise to signal plus noise distributions for either the number of defectives or the number of nondefectives in the sample.

Statistical literature contains a variety of formulations of the notion of "almost binomial" data. Among the most widely studied relatives of the binomial distribution are: truncated binomial distributions, for which moment estimators were obtained by Rider (1955) and Shah (1966) and for which maximum likelihood estimates were derived by Finney (1949); mixtures of binomial distributions studied by Blischke (1964); censored samples from binomial distributions, for which Blight (1970) developed useful theory. A formulation similar to that studied here is the model for misclassification of Bernoulli data studied by Bryson (1965). Models for binomial signals in discrete noise may be viewed as misclassification models in which the counting mechanism systematically overcounts binomial data.

In Section II, we establish a characterization result for convoluted binomial distributions in terms of differential equations satisfied by their probability mass functions. We show in Section III that the

characterization result may be fruitfully applied to maximum likelihood estimation of the binomial parameter for a fairly broad class of signal plus noise models.

II. CONVOLUTED BINOMIAL DISTRIBUTIONS

Katz (1946) characterized a trilogy of discrete distributions in terms of the difference equation

$$(x+1)f(x+1) = (a+bx)f(x) \tag{2.1}$$

satisfied by the probability mass function f. He showed that the difference equation characterized Poisson distributions among distributions with equal mean and variance, while it characterized binomial (negative binomial) distributions among distributions with mean smaller (larger) than its variance. The characterization result we obtain below is the third of a similar family of characterizations -- these in terms of differential equations satisfied by the probability mass functions. These latter results have the form

$$\frac{\partial}{\partial\theta} P_\theta(x) = P_\theta(x-1) - P_\theta(x) \tag{2.2}$$

where P_θ represents the probability mass function of a variable in a given parametric family. The result established here characterizes convoluted binomial distributions by equations of the form (2.2). The complementary result for convoluted Poisson distributions is given in Samaniego (1976), and for convoluted Pascal (negative binomial) distributions in Hannon and Samaniego (1977).

Differential equations of the form (2.2) have been investigated earlier. Boswell and Patil (1973) obtained characterizations of the Poisson, binomial and negative binomial distributions in terms of these equations. Our work has extended their results to convolutions of discrete distributions, and utilized these extensions in maximum likelihood estimation for signal plus noise distributions. Our characterization of convoluted binomial distributions is as follows.

<u>Theorem</u>. Let $\{X_{N,p}\}$ be a family of nonnegative integer valued random variables indexed by $p \in (0,1)$ and $N \in \{1,2,\cdots\}$. Let Z be a discrete variable whose distribution does not depend on p. Define $X_{0,p} \equiv Z$, and suppose that, for every N, $X_{N,p} \xrightarrow{\mathcal{L}} Z$ as $p \to 0$. Then, $P(X=k|N,p)$ is differentiable in p and

$$\frac{\partial}{\partial p} P(X=k|N,p) = N[P(X=k-1|N-1,p) - P(X=k|N-1,p)] \tag{2.3}$$

for all $N \geq 1$, $p \in (0,1)$ and $k \geq 0$ if, and only if, the distribution of $X_{N,p}$ is a convolution of the binomial distribution with parameters N and p and the distribution of the random variable Z.

<u>Proof</u>: We denote the binomial distribution with parameters N and p by $\mathcal{B}(N,p)$. It is easy to verify that if the family $\{X_{N,p}\}$ consists precisely of sums $Y_{N,p} + Z$, where Y and Z are independent and $Y \sim \mathcal{B}(N,p)$, then the probability mass function of $X_{N,p}$ satisfies (2.3) for all appropriate N, p and k. We prove the converse by induction on N. Suppose (2.3) is satisfied for $N \geq 1$, $p \in (0,1)$ and $k \geq 0$. Consider $N = 1$. Let k be a fixed but arbitrary nonnegative integer. We have by (2.3) that

$$\frac{\partial}{\partial p} P(X=k|1,p) = P(X=k-1|0,p) - P(X=k|0,p)$$

$$= P(Z=k-1) - P(Z=k)$$

The general solution of this differential equation is

$$P(X=k|1,p) = p(P(Z=k-1) - P(Z=k)) + c_k$$

where c_k is a constant independent of p. The boundary conditon $X_{1,p} \xrightarrow{\mathcal{L}} Z$ as $p \to 0$ identifies c_k as $P(Z=k)$, so that

$$P(X=k|1,p) = pP(Z=k-1) +(1-p)P(Z=k).$$

Thus, $X_{1,p}$ has the distribution of $Y_{1,p} + Z$, where Y and Z are independent and $Y \sim B(1,p)$. We now assume that for $N < N^*$, $X_{N,p}$ has the distribution of the sum $Y_{N,p} + Z$, where Y and Z are independent and $Y \sim B(N,p)$. For $N=N^*$, we have for each fixed k that

$$\frac{\partial}{\partial p} P(X=k|N^*,p) = N^*[P(X=k-1|N^*-1,p) - P(X=k|N^*-1,p)] \quad (2.4)$$

By the induction hypothesis, we may replace the probabilities on the right hand side of (2.4) by the probabilities of the appropriate convoluted binomial distribution. Equating the indefinite integrals of both sides of (2.4) yields the general solution

$$P(X=k|N^*,p) = \sum_{i=0}^{\min(k,N^*)} \binom{N^*}{i} p^i (1-p)^{N^*-i} P(Z=k-i) + c_k$$

where c_k is a constant independent of p. Invoking the fact that $X_{N^*,p} \xrightarrow{\mathcal{L}} Z$ as $p \to 0$, we have $c_k = 0$. Thus, $X_{N^*,p}$ has the distribution

of $Y_{N^*,p} + Z$ where Y and Z are independent and $Y \sim \mathcal{B}(N^*,p)$, completing the proof.

Convoluted binomial distributions characterized above may be viewed as mixtures of shifted binomial distributions. If $X \sim \mathcal{B}_a(N,p)$ is understood to mean that $X - a \sim \mathcal{B}(N,p)$, then the convolution of $\mathcal{B}(N,p)$ with the distribution $\{p_i\}_{i=0}^{\infty}$ may be represented as the mixture $\Sigma\, p_i B_i(N,p)$. The method of moments has often proven tractable in estimating the parameters of mixtures of distributions, and might be pursued in dealing with convoluted binomial distributions. We choose instead to examine maximum likelihood estimation in the next section, where we are able to directly apply the system of differential equations (2.3) satisfied by convoluted binomial distributions.

III. MAXIMUM LIKELIHOOD ESTIMATION

Estimation of the parameters of a convolution by the method of maximum likelihood is in general a formidable analytic problem. For convolutions of discrete distributions, the likelihood function for a random sample is a product of somewhat unmanageable sums, and a straightforward approach to maximization seems unpromising. Sclove and Van Ryzin (1969) proposed method of moments estimators for parameters of convolutions, showing this approach to be tractable in many examples. Samaniego (1976) showed that maximum likelihood estimation was feasible for certain convoluted Poisson distributions. We show here that a similar analysis can be applied to convoluted binomial distributions.

Let $X_1,\ldots,X_n$ be a sample of size n from a convoluted binomial distribution with known noise distribution. Denote the likelihood function for this sample by

$$L(x_1,x_2,\ldots,x_n,p) = \prod_{i=1}^{n} P(X_i=x_i|N,p)$$

The characterization result established in Section II implies that the likelihood equation $\frac{\partial}{\partial p} \ln L = 0$ may be written as

$$\sum_{i=1}^{n} \frac{P(X_i=x_i-1|N-1,p) - P(X_i=x_i|N-1,p)}{P(X_i=x_i|N,p)} = 0. \qquad (3.1)$$

We define below, for a parametric family of probability distributions, the notion of parametric monotone ratio in a given parameter.

<u>Definition</u>. Let $\{P_{p,\alpha}, p \in \mathcal{P}, \alpha \in \mathcal{A}\}$ be a family of discrete distributions indexed by a real parameter p and a (possibly degenerate) parameter α. The family is said to have parametric monotone decreasing ratio (PMDR) in p if, for each fixed α and for each $a < b$ for which a and b receive positive mass under $P_{p,\alpha}$ for every $p \in \mathcal{P}$, the ratio

$$\frac{P_{p,\alpha}(X=a)}{P_{p,\alpha}(X=b)}$$

is decreasing in p.

The relevance of PMDR to maximum likelihood estimation for convoluted binomial distributions is seen in the following lemma.

Lemma 1. Let $X_1,\ldots,X_n$ be a random sample from a convoluted binomial distribution having PMDR in the parameter p. Then $\frac{\partial}{\partial p} \ell n\, L$ is a decreasing function of p.

Proof: Since $\frac{\partial}{\partial p} \ell n\, L = \sum_{i=1}^{n} \frac{\partial}{\partial p} \ell n\, P(X_i=x_i|N,p)$ we need only show that $\frac{\partial}{\partial p} \ell n\, P$ is decreasing in p. We assume that x is such that $P(X=x|p,N) > 0$ for some (and thus for all) $p \in (0,1)$. Note that

$$\frac{\partial}{\partial p} \ell n\, P(X=x|N,p)$$

$$= N \frac{P(X=x-1|N-1,p) - P(X=x|N-1,p)}{P(X=x|N,p)}$$

$$= N \frac{P(X=x-1|N-1,p) - P(X=x|N-1,p)}{pP(X=x-1|N-1,p) + (1-p)P(X=x|N-1,p)}$$

$$= N \frac{P(X=x-1|N-1,p) - P(X=x|N-1,p)}{p(P(X=x-1|N-1,p) - P(X=x|N-1,p)) + P(X=x|N-1,p)} \,. \tag{3.2}$$

For a convoluted binomial distribution with $P(X=x|N,p) > 0$, the terms $P(X=x|N-1,p)$ and $P(X=x-1|N-1,p)$ cannot both be identically zero in p. Moreover, because of PMDR, their difference can be zero for at most one value of p. If the numerator is positive (negative) for all $p \in (0,1)$, then (3.2) is a decreasing positive (negative) function of p. This includes the cases where either $P(X=x|N-1,p) = 0$ or $P(X=x-1|N-1,p) = 0$ for which (3.2) is equal to $1/p$ and $-1/(1-p)$ respectively. Assume that $P(X=x|N-1,p) > 0$ and suppose the numerator of (3.2) is zero at $p_0 \in (0,1)$. The ratio (3.2) is of the form

$$\frac{h(p)}{g(p) + f(p)} \, .$$

For $p < p_0$, $h(p)/g(p) = 1/p$ and $h(p)/f(p)$ is nondecreasing, so that (3.2) is positive and decreasing for $p < p_0$. The ratio (3.2) equals zero at $p = p_0$. For $p > p_0$, we find that (3.2) is negative and decreasing, so that (3.2) is decreasing for $p \in (0,1)$. This shows that $\frac{\partial}{\partial p} \ell n\, P$ is a decreasing function of p, completing the proof.

Suppose a random sample is taken from a convoluted binomial distribution with PMDR in p, and suppose that the noise distribution is known. Lemma 1 implies that the likelihood equation $\frac{\partial}{\partial p} \ell n\, L = 0$ has at most one root. Thus, the maximum likelihood estimate of p is zero, one or the unique solution of the likelihood equation. The MLE may easily be obtained numerically for such problems, since values of $\frac{\partial}{\partial p} \ell n\, L$ above and below zero may be used to approximate the MLE to any desired degree of accuracy.

We consider below examples of convoluted binomial distributions with PMDR. In Example 1, we will invoke the following result.

Lemma 2. Let $\{f_i\}$ be a sequence of positive functions of a real variable y defined on an interval of the real line. Let $\{a_i\}$, $\{b_i\}$ be sequences of positive real numbers, with $a_i < a_{i+1}$ for all i. If $f_i(y)/f_j(y)$ is decreasing in y whenever $i < j$, then every function of the form

$$H_n(y) = \frac{\sum_{i=1}^{n} b_{k_i} f_{k_i}(y)}{\sum_{i=1}^{n} a_{k_i} b_{k_i} f_{k_i}(y)}, \qquad k_1 < k_2 < \cdots < k_n,$$

for $n \geq 2$, is decreasing in y.

Proof: To start an induction argument, we consider $n = 2$.

$$H_2(y) = \frac{b_{k_1} f_{k_1}(y) + b_{k_2} f_{k_2}(y)}{a_{k_1} b_{k_1} f_{k_1}(y) + a_{k_2} b_{k_2} f_{k_2}(y)}$$

$$= \frac{b_{k_1} f_{k_1}(y) + b_{k_2} f_{k_2}(y)}{a_{k_1}(b_{k_1} f_{k_1}(y) + b_{k_2} f_{k_2}(y)) + (a_{k_2} - a_{k_1}) b_{k_2} f_{k_2}(y)}$$

which is of the form

$$\frac{h(y)}{ch(y) + g(y)} .$$

Since $h(y)/g(y)$ is decreasing by hypothesis, we have that $H_2(y)$ is decreasing in y. Suppose the lemma holds for n. Then

$$H_{n+1}(y) = \frac{\sum_{i=1}^{n+1} b_{k_i} f_{k_i}(y)}{\sum_{i=1}^{n+1} a_{k_i} b_{k_i} f_{k_i}(y)} , \qquad k_1 < \cdots < k_{n+1},$$

$$= \frac{\sum_{i=1}^{n+1} b_{k_i} f_{k_i}(y)}{a_{k_1}\left(\sum_{i=1}^{n+1} b_{k_i} f_{k_i}(y)\right) + \sum_{i=2}^{n+1} (a_{k_i} - a_{k_1}) b_{k_i} f_{k_i}(y)}$$

which is decreasing in y if

$$\frac{\sum_{i=1}^{n+1} b_{k_i} f_{k_i}(y)}{\sum_{i=2}^{n+1} (a_{k_i} - a_{k_1}) b_{k_i} f_{k_i}(y)} \tag{3.3}$$

is decreasing.

But by hypothesis,

$$\frac{f_{k_1}(y)}{\sum_{i=2}^{n+1} (a_{k_i} - a_{k_1}) b_{k_i} f_{k_i}(y)}$$

is decreasing in y, so the ratio (3.3) is decreasing if

$$\frac{\sum_{i=2}^{n+1} b_{k_i} f_{k_i}(y)}{\sum_{i=2}^{n+1} (a_{k_i} - a_{k_1}) b_{k_i} f_{k_i}(y)}$$

is decreasing. This latter ratio has the form of $H_n(y)$ and is decreasing in y by the induction hypothesis. Thus, the proof is complete.

Example 1. Let $Y \sim \mathcal{B}(N,p)$, and $Z \sim \mathcal{P}(\theta)$, i.e., Z is a Poisson random variable with parameter θ. Let Y, Z be independent and let $X = Y + Z$. One can easily check that the ratio $P(X=0)/P(X=1)$ is decreasing in p. For $x \geq 2$,

$$\frac{P(X=x-1)}{P(X=x)} = \frac{\sum_{i=0}^{x-1} \frac{\theta^{x-i-1}}{(x-i-1)!} P_p(Y=i)}{\sum_{i=0}^{x} \frac{\theta^{x-i}}{(x-i)!} P_p(Y=i)}$$

which is decreasing in p if the ratio (3.4) below is decreasing

$$\frac{\sum_{i=0}^{x-1} \frac{\theta^{x-i-1}}{(x-i-1)!} P_p(Y=i)}{\sum_{i=0}^{x-1} \frac{\theta^{x-i}}{(x-i)!} P_p(Y=i)}, \tag{3.4}$$

since when $P_p(Y=x) > 0$, $P_p(Y=i)/P_p(Y=x)$ is decreasing in p for all $i < x$ for which $P_p(Y=i) > 0$. The function (3.4) satisfies the hypotheses of Lemma 2, being of the form $H_{n+1}(p)$, with

$$f_i(p) = P_p(Y=i)$$

$$b_i = \frac{\theta^{x-i-1}}{(x-i-1)!}$$

and

$$a_i = \frac{\theta}{x-i}$$

for $k_i = i = 0,1,\ldots,n$, where $n = \min(N,x-1)$. It follows that the function (3.4), and consequently the ratio $P(X=x-1)/P(X=x)$, is decreasing in p. Thus, the binomial-Poisson convolution has PMDR in the parameter p.

Example 2. Let Y be a random variable whose distribution belongs to a family of distributions with PMDR in the parameter p. Suppose the support of Y is equal to a set of consecutive positive integers. If $Z \sim \mathcal{B}(1,\pi)$ and Y and Z are independent, we show that the distribution of $X = Y + Z$ has PMDR in p. It is easy to check that $P(X=a)/P(X=a+1)$ is decreasing, where a is the smallest integer that Y takes on with positive probability. For any $y > a$ for which $P(Y=y) > 0$,

$$\frac{P(X=y-1)}{P(X=y)} = \frac{\pi P(Y=y-2) + (1-\pi)P(Y=y-1)}{\pi P(Y=y-1) + (1-\pi)P(Y=y)}$$

$$= \frac{\pi \cdot \dfrac{P(Y=y-2)}{P(Y=y-1)} + (1-\pi)}{\pi + (1-\pi)\dfrac{P(Y=y)}{P(Y=y-1)}}$$

which is decreasing in p. If Y has finite support and $N = \text{lub}\{\text{Support of } Y\}$, then

$$\frac{P(X=N)}{P(X=N+1)} = \frac{\pi P(Y=N-1) + (1-\pi)P(Y=N)}{\pi P(Y=N)}$$

which is also decreasing in p. Thus $X = Y + Z$ has a distribution with PMDR. Since the support of X is again equal to a set of consecutive positive integers, it follows that the distribution of $X = Y + Z$, with Y, Z independent and $Z \sim \mathcal{B}(M,\pi)$ has PMDR in p. Moreover, since the binomial distribution itself has support on consecutive positive integers, we conclude that convoluted binomial distributions $\mathcal{B}(N,p) * \mathcal{B}(M,\pi)$ have PMDR in p.

In Example 2, we require the support of a random variable Y to consist of a set of consecutive integers for convolution with a Bernoulli variable to be PMDR preserving. The necessity of this condition is apparent in the following example.

<u>Example 3</u>. Let Y be a variable taking values on even integers according to binomial probabilities as follows:

$$P(Y=2i) = \binom{N}{i} p^{i}(1-p)^{N-i} \qquad i=0,1,\ldots,N.$$

It is obvious that the distribution of Y has PMDR. Let $Z \sim \mathcal{B}(1,\pi)$ be independent of Y, and let $X = Y + Z$. For even x,

$$\frac{P(x-1)}{P(x)} = \frac{\pi P(Y=x-2)}{(1-\pi)P(Y=x)}$$

which is decreasing in p. If, however, x is odd,

$$\frac{P(x-1)}{P(x)} = \frac{(1-\pi)P(Y=x-1)}{\pi P(Y=x-1)} = \frac{1-\pi}{\pi} .$$

Thus, the ratio $P(x-1)/P(x)$ is a constant independent of p.

In Examples 1 and 2 given above, PMDR in the parameter p guarantees that the likelihood equation has at most one solution. As we shall see, there are convoluted binomial distributions without PMDR, but satisfying a weaker monotonicity property, for which we are able to demonstrate that the likelihood equation has at most one root.

Definition. Let $\{P_{p,\alpha}, p \in \mathcal{P}, \alpha \in \mathcal{A}\}$ be a family of discrete distributions indexed by a real valued parameter p and a (possibly degenerate) parameter α. The family is said to have weak parametric monotone decreasing ratio in p (WPMDR) if, for each fixed $\alpha \in \mathcal{A}$ and for each $a < b$ for which a and b receive positive mass under $P_{p,\alpha}$ for every $p \in \mathcal{P}$, the ratio

$$\frac{P_{p,\alpha}(X=a)}{P_{p,\alpha}(X=b)}$$

is either decreasing in p or is equal to a constant independent of p.

Lemma 3. Let $X_1,\ldots,X_n$ be a random sample from a convoluted binomial distribution $P_{N,p}$ with WPMDR in the parameter p. If at least one observation x_i is such that

$$P(X=x_i-1|N-1,p)/P(X=x_i|N-1,p) \tag{3.5}$$

is not identically 1 in p, then $\frac{\partial}{\partial p}\ \ell n\ L$ is a decreasing function of p.

Proof: For every observation x_i for which the ratio (3.5) is identically 1, we have $\frac{\partial}{\partial p}\ \ell n\ P(X=x_i|N,p) = 0$ by (2.3). For any observation x_i for which the ratio (3.5) is decreasing, $\frac{\partial}{\partial p}\ \ell n\ P(X=x_i|N,p)$ is decreasing by the argument used in proving Lemma 1. It remains to show that this is also true for observations x_i for which the ratio (3.5) is identically equal to $c \neq 1$. We have

$$\frac{\partial}{\partial p}\ \ell n\ P(X=x_i|N,p) = \frac{P(X=x_i-1|N-1,p) - P(X=x_i|N-1,p)}{p(P(X=x_i-1|N-1,p) - P(X=x_i|N-1,p)) + P(X=x_i|N-1,p)}\ .$$

This function has the form

$$\frac{h(p)}{g(p) + f(p)}$$

where $h(p)/g(p) \equiv 1/p$ and $h(p)/f(p) \equiv c - 1$. Thus $\frac{\partial}{\partial p}\ \ell n\ P(X=x_i|N,p)$ is decreasing in p. Since $\frac{\partial}{\partial p}\ \ell n\ L$ is the sum of functions each of which is decreasing in p or identically zero, and at least one of these functions is decreasing, we have that $\frac{\partial}{\partial p}\ \ell n\ L$ is decreasing in p.

We now consider a pair of examples of convoluted binomial distributions with weak PMDR.

Example 4. Let Y be a nonnegative integer valued random variable whose distribution belongs to a one-parameter family of distributions with WPMDR in the parameter p. If Z is independently distributed according to a geometric distribution $G(\theta)$ with probability mass function

$$f(Z=z) = (1-\theta)\theta^{z} \qquad z=0,1,2,\cdots ,$$

we show that the distribution of $X = Y + Z$ has WPMDR in p. Suppose $x - 1$ and x receive positive mass from the distribution of X. If $P(Y=x) = 0$,

$$\frac{P(X=x-1)}{P(X=x)} = \frac{\sum_{i=0}^{x-1} (1-\theta)\theta^{i} P_p(Y=x-i-1)}{\sum_{i=0}^{x} (1-\theta)\theta^{i} P_p(Y=x-i)}$$

is equal to $1/\theta$. If $P_p(Y=x) > 0$, this ratio may be written in the form

$$\frac{h(p)}{g(p) + f(p)} ,$$

where $h(p)/g(p) = 1/\theta$ and $h(p)/f(p)$ is either constant or a decreasing function of p. Thus, for all appropriate x, the ratio $P(X=x-1)/P(X=x)$ is either constant or decreasing in p, establishing WPMDR for the distribution of X. Since the binomial distribution $\mathcal{B}(N,p)$ has WPMDR, we have that the convolution $\mathcal{B}(N,p) * G(\theta)$ has WPMDR. It is useful to note that for $X \sim \mathcal{B}(N,p) * G(\theta)$, the ratio $P(x-1)/P(x)$ is decreasing in p for $0 \leq x \leq N$, and is equal to $1/\theta$ for $x > N$. Since this ratio is never identically 1, the function $\frac{\partial}{\partial p} \ln L$ is decreasing in p for samples

from $\mathcal{B}(N,p) * G(\theta)$.

Consider the convolution $\mathcal{B}(N,p) * G(\theta) * G(\theta)$, or equivalently the convolution $\mathcal{B}(N,p) * n\mathcal{B}(2,\theta)$ where $n\mathcal{B}(r,\theta)$ represents the negative binomial distribution with parameters r and θ. If $Y \sim \mathcal{B}(N,p) * G(\theta)$ and $Z \sim G(\theta)$, independent of Y, and if $X = Y + Z$, then for any $x > 0$

$$\frac{P(x-1)}{P(x)} = \frac{\sum_{i=0}^{x-1} \theta^i P(Y=x-i-1)}{\theta \sum_{i=0}^{x-1} \theta^i P(Y=x-i-1) + P_p(Y=x)} .$$

Because $P_p(Y=x)$ is positive for all p, this ratio is decreasing in p. Thus the family $\mathcal{B}(N,p) * n\mathcal{B}(2,\theta)$ has PMDR in p. The same argument leads one to conclude that $\mathcal{B}(N,p) * n\mathcal{B}(r,\theta)$ has PMDR in p for any $r \geq 2$.

Example 5. Let $Y \sim \mathcal{B}(N,p)$, and let Z be a random variable independent of Y with a uniform distribution on the integers $0,1,\ldots,k$, that is, with probability mass function

$$P(Z=z) = \frac{1}{k+1}, \qquad z=0,1,\ldots,k.$$

Suppose $k > N$. Then the ratio

$$\frac{P(x-1)}{P(x)} = \begin{cases} \dfrac{\sum_{i=0}^{x-1} P_p(Y=i)}{\sum_{i=0}^{x} P_p(Y=i)} & \text{if } x \leq N \\[2ex] 1 & \text{if } N < x \leq k \\[2ex] \dfrac{\sum_{i=x-k-1}^{N} P_p(Y=i)}{\sum_{i=x-k}^{N} P_p(Y=i)} & \text{if } k < x \leq N + k \end{cases}$$

It is therefore clear that the distribution of X has WPMDR.

Suppose a random sample of size n is taken from the convolution $\mathcal{B}(N,p) * \mathcal{U}[0,1,\ldots,k]$, with $k \geq N$. The distribution $\mathcal{B}(N-1,p) * \mathcal{U}[0,1,\ldots,k]$ has WPMDR in p. If at least one observation x_i fails to satisfy the inequality $N \leq x_i \leq k$, we have by Lemma 3 that $\frac{\partial}{\partial p} \ell n\, L$ is decreasing in p. If all $x_i \leq k$ with at least one $x_i < N$, the MLE is $\hat{p} = 0$. If all $x_i \geq N$ with at least one $x_i > k$, the MLE is $\hat{p} = 1$. In all other cases in which observations x_i smaller than N or larger than k exist, the MLE is the unique solution of the likelihood equation. When all x_i satisfy $N \leq x_i \leq k$, the likelihood is equal to the constant $(1/k+1)^n$, independent of p, and any $p \in [0,1]$ is an MLE.

For completeness, we note that the convolution $\mathcal{B}(N,p) * \mathcal{U}[0,\ldots,k]$ for $k \leq N$ has PMDR in p. Thus, if a sample is taken from $\mathcal{B}(N,p) * \mathcal{U}[0,\ldots,k]$ with $k < N$, we have by Lemma 1 that $\frac{\partial}{\partial p} \ell n\, L$ is decreasing in p and the MLE is zero, one or the unique solution of the likelihood equation.

The examples given in this section serve to demonstrate that maximum likelihood estimation of the binomial parameter is easily accomplished for a fairly broad class of convoluted binomial distributions. In addition to the distributions examined in these examples, multiple convolutions of $\mathcal{B}(N,p)$ with two or more other families, for example, the three-fold convolution $\mathcal{B}(N,p) * \mathcal{P}(\theta) * \mathcal{NB}(r,\pi)$, can be shown to have PMDR in p. In all such cases, with all parameters save the binomial parameter p known, the maximum likelihood estimate of p is easily found. It is possible to construct convoluted binomial distributions for which the equation $\frac{\partial}{\partial p} \ell n\, L = 0$ has any number of solutions. Thus, maximum likelihood

estimation can be unfeasible for such models even by numerical methods. When parametric monotonicity holds, the estimation problem is entirely straightforward.

REFERENCES

Blight, B. J. N. (1970). Estimation from a censored sample for the exponential family. Biometrika 57, 389-95.

Blischke, W. R. (1964). Estimating the parameters of mixtures of binomial distributions. Journal of the American Statistical Association 59, 510-28.

Boswell, M. T. and Patil, G. P. (1973). Characterization of certain discrete distributions by differential equations with respect to their parameters. Australian Journal of Statistics 15, 128-31.

Bryson, M. R. (1965). Errors of classification in a binomial population. Journal of the American Statistical Association 60, 217-24.

Crochiere, R. E., Webber, S. A. and Flanagan, J. L. (1976). Digital coding of speech in sub-bands. Bell System Technical Journal 55, 875-904.

Finney, D. J. (1949). The truncated binomial distribution. Annals of Eugenics 14, 319-28.

Gaffey, W. R. (1959). A consistent estimator of a component of a convolution. Annals of Mathematical Statistics 30, 198-205.

Greenstein, L. J. (1974). Spectrum of a binary signal block coded for D.C. suppression. Bell System Technical Journal 53, 1103-26.

Hannon, G. G. and Samaniego, F. J. (1977). A generalization of a theorem of Boswell and Patil. Technical Report, Department of Mathematics, University of California, Davis.

Katz, L. (1946). On a class of functions defined by the difference equation $(x+1)f(x+1) = (a+bx)f(x)$. (abstract) Annals of Mathematical Statistics 17, 501.

Samaniego, F. J. (1976). A characterization of convoluted Poisson distributions with applications to estimation. Journal of the American Statistical Association 71, 475-79.

Sclove, S. L. and Van Ryzin, J. (1969). Estimating the parameters of a *convolution*. Journal of the Royal Statistical Society B **31**, 181-91.

Shah, S. M. (1966). On estimating the parameter of a doubly truncated binomial distribution. Journal of the American Statistical Association **61**, 259-63.

Woodward, P. M. (1953). Probability and Information Theory with Applications to Radar, Oxford: Pergamon Press.